

WERNER KALLMEYER

Möglichkeiten der maschinellen Unterstützung bei der Arbeit mit Interaktionskorpora

Abstract

In diesem Beitrag sollen anhand von Materialien aus Gesprächskorpora des IDS Schwierigkeiten und Möglichkeiten der maschinellen Recherche vorgeführt werden. Grundlage dafür sind Gesprächstranskripte, die in digitaler Form vorliegen und in einem System mit Rechercheprozeduren zugreifbar sind. Mit diesem Ziel wird auf Rechercheverfahren zurückgegriffen, die in den 1990er Jahren in einem Projekt SERGES im IDS als Anwendung der COSMAS-Technologie auf Gesprächskorpora entwickelt wurden. Die hier gegebenen Recherchemöglichkeiten werden an einem Auswahlkorpus von Gesprächstranskripten mit einem Gesamtumfang von 87.629 laufenden Wörtern versuchsweise angewendet und in ihren Beschränkungen und ihrer Fruchtbarkeit für explorative Untersuchungen betrachtet. Damit soll ein Beitrag zur Klärung der Frage geleistet werden, welche Recherchemöglichkeiten aus einer gesprächsanalytischen Perspektive vorstellbar und erwünscht sind und insofern bei der weiteren korpus technologischen Entwicklung berücksichtigt werden sollten.

1. Zu den Voraussetzungen maschineller Recherchen in Interaktionskorpora

Interaktionskorpora enthalten: Ton- bzw. Video-Dokumente von sprachlichen Interaktionen in ihrem natürlichen Vorkommenszusammenhang, Dokumentation der Situation und zur Einbettung der jeweiligen Interaktionszene in übergreifende Zusammenhänge, Transkriptionen der sprachlichen und manchmal auch der nicht-sprachlichen Kommunikationsvorgänge. Solche Korpora entstanden im Zusammenhang mit der Entwicklung der Konversationsanalyse seit den 60er Jahren, in Deutschland seit dem Beginn der 1970er Jahre. Es gab und gibt vielerorts kleinere Korpora von Interaktionsmaterialien als Ergebnis von Forschungsprojekten. Unter dem Einfluss der digitalen Aufnahme-, Präsentations- und Archivierungstechnik wird auch der weitere Aufbau von Interaktionskorpora inzwischen von der Erwartung geleitet, dass die Aufnahmen ebenso wie die Transkriptionen und sonstigen schriftlichen Materialien in digitaler Form vorliegen, über eine Verwaltungskomponente auffindbar und zugreifbar sind und dass nach Möglichkeit Aufnahmen und Transkriptionen über Zeitindices miteinander verknüpft sind (Alignment), so dass zu ausgewählten Transkriptionsstücken die Ton/Bild-Abschnitte präsentiert werden können.

Größere digitale Interaktionskorpora mit einer solchen Datenbankstruktur existieren gegenwärtig aber erst im Ansatz. In dieser Hinsicht liegen Interaktionskorpora weit hinter den Korpora geschriebener Sprache zurück. Im British National Corpus wurde als ein Kriterium für ein nationales Referenzkorpus festgelegt, dass 10 % mündliche Sprache enthalten sein müsste. Diese Anforderung wird von der Vorstellung geleitet, dass alle wesentlichen Merkmale einer Sprache von schriftlichen Texten abzulesen seien mit Ausnahme der phonetisch-phonologischen Aspekte. In diesem Sinne entsprechen die 10 % gesprochener Sprache einer Minimallösung. Aber auch diese 10 % werden nicht erreicht. Korpora gesprochener Sprache wachsen relativ langsam. Das gilt insbesondere für Interaktionskorpora.

Die beim Aufbau von Korpora gesprochener Sprache auftretenden Probleme von der Daten-Beschaffung über die Transkriptionsweise, die Dokumentation von nicht-sprachlichen Eigenschaften von Interaktionsereignissen bis zur Frage der mehrstufigen Annotation sind in den letzten Jahren im Zusammenhang mit bedeutenden Initiativen des Korpusaufbaus vielfältig reflektiert worden.¹ Diese Diskussion soll hier nicht noch einmal nachgezeichnet werden. Vielmehr konzentriert sich dieser Beitrag auf die antizipierende Entwicklung von Recherchemöglichkeiten in Interaktionskorpora.

Der große Herstellungsaufwand von Interaktionskorpora lohnt sich im Kern nur für Interaktionsanalysen, also Untersuchungen, die sprachliche Phänomene relativ zu Sprechern, Sprecherwechseln und sequenziellen Positionen innerhalb von Redebeiträgen und in Folgen von Redebeiträgen erfassen. Trotzdem ist es natürlich sinnvoll, die Interaktionskorpora so einzurichten, dass sie auch für andere Untersuchungsinteressen, die nicht unmittelbar auf die Interaktionsstruktur rekurrieren, nutzbar sind. Langfristig ist das auch für die Interessen der Interaktionsanalyse von großem Interesse. Am besten realisierbar sind zurzeit „einfache“ Dinge wie Wortsuche. Möglich ist damit die Untersuchung von „mündlicher Sprache“ anhand von Wortformen (und ggf. Wörtern) und Kollokationen. Je realistischer die schriftliche Wiedergabe der Lautform ist, umso schwieriger wird allerdings die Lemmatisierung. Diese ist bei einer von der Orthographie abweichenden Schreibweise, z. B. der literarischen Umschrift mit der Wiedergabe von Schnellsprech-Formen, Verschleifungen, dialektalen Prägungen usw. nicht unmittelbar möglich, sondern nur über eine Normalisierung der Verschriftung.

Lemmatisierung vorausgesetzt kann man anhand von Interaktionskorpora (zusammen mit Korpora monologischer Daten gesprochener Sprache) den Wortschatz des gesprochenen Deutsch dokumentieren und sein Vorkommen in Textsorten / Interaktionstypen und bezogen auf Sprecherkategorien (also z. B. Männer / Frauen, ggf. Altersklassen, soziale Schicht usw.) untersuchen.

¹ Vgl. u. a. Leech/Myers/Thomas 1955; Pusch/Raible (2002); Schwitalla/Wegstein (2005).

Diese Art der Auswertung hängt ab von den verfügbaren Metadaten (z. B. für eine Sortierung nach Interaktionstypen, Formalität und Informalität, Geschlecht und Alter der Sprecher und nach der Entstehungszeit). Das Problem ist hier eher die Verfügbarkeit der Daten in ausreichender Menge als die Recherchestrategien. Mögliche Fragen sind weiter die nach der Anzahl der laufenden Wörter in einer Interaktion, bei einzelnen Sprechern und ggf. auch noch innerhalb von Gesprächsphasen (bei einer entsprechenden Segmentierung des Gesprächsverlaufs).

Im Folgenden soll erläutert werden, was man mit einem Recherche-System, das über Suchoperatoren für Wörter und Interaktionseigenschaften verfügt wie das experimentelle System COSMAS II für Gesprächsdaten, im Rahmen von qualitativ ausgerichteten Untersuchungen tun kann.² Es geht dabei darum, begrenzte Suchmöglichkeiten heuristisch zu nutzen, Grenzen zu reflektieren und Recherchestrategien weiter zu entwickeln und Entwicklungsbedarf zu präzisieren für die Recherche von komplexen Phänomenen, die gegenwärtig kaum erfassbar sind. Da das System bislang nur in einer Versuchsversion existiert und nur ein begrenztes Korpus angeschlossen ist, wurden für die Zwecke dieser Demonstration und Reflexion von Möglichkeiten und Grenzen die Suchvorgänge mit anderen, einfacheren Werkzeugen wie Wortsuche in WORD ansatzweise simuliert und sorgfältig von Hand kontrolliert. Es handelt sich also bei dem, was hier vorgestellt wird, nicht um eine Recherche in einem voll funktionstüchtigen System, sondern um die Simulation der Arbeit mit Recherche-Operatoren, für die anhand von einem Entwicklungskorpus bereits Erfahrungen gesammelt wurden, teilweise auch für die hier interessierenden Recherchen.³

2. Das Zielobjekt

Im Zentrum des Interesses steht ein grammatisch-pragmatisches Muster, in dem zwei Äußerungseinheiten miteinander verknüpft werden, zwischen denen eine Beziehung der Opposition bzw. partiellen Inkompatibilität besteht: JA-ABER. Aus früheren gesprächsanalytischen Studien zu derartigen Phänomenen hat sich als Hypothese ergeben, dass es sich um ein relativ komplexes, formal und funktional bestimmtes Muster handelt, das eine verfestigte gesprächsrhetorische Lösung für bestimmte Anforderungen des Formulierens im Gespräch darstellt. Die Grundaufgabe kann man angeben als: die Relevanzen anderer berücksichtigen und eigene Relevanzen verfolgen.⁴ Das Muster hat folgende Eigenschaften:

² Einige der im Experimentalsystem enthaltenen Suchmöglichkeiten sind auch der DGD implementiert, allerdings nicht alle; und zudem ist die Menge der Gesprächstranskripte in der DGD nach wie vor sehr begrenzt.

³ Das gilt z. B. für die Suche nach Verwendungen des Konnektors *aber*; vgl. Kallmeyer/Schütte 2005.

⁴ Die Beschäftigung mit solchen Konstruktionsmustern war Gegenstand eines Projekts

- Formuliert wird eine komplexe Äußerung mit zwei Teilen. Ein Teil allein ist hinsichtlich der Äußerungskonstitution unvollständig ohne den anderen Teil. Der Normalfall ist, dass beide Teile vom selben Sprecher produziert werden.
- Der erste Teil drückt eine Berücksichtigung und / oder Bestätigung einer Position aus und der zweite Teil eine Zurückweisung bzw. eine mit der ersten nicht kompatible andere Position.
- Der erste Teil bezieht sich auf eine fremde Position, insbesondere auf eine Position eines Interaktionspartners, aber auch auf Positionen Dritter oder auf allgemein mögliche und nicht personal zugeschriebene Positionen. Der zweite Teil bezieht sich auf eine Position des aktuellen Sprechers.
- Es besteht ein Relevanzgefälle zwischen dem ersten und dem zweiten Teil in dem Sinne, dass der erste Teil in der Relevanz rückgestuft wird und der zweite Teil hochgestuft.
- Für die Markierung des zweiten Teils werden adversative Konnektoren wie *aber*, *nur*, *bloß* verwendet, insgesamt eine relativ begrenzte Menge. Die Markierung des ersten Teils ist wesentlich variabler. Neben *zwar* gibt es vor allem *ja* und andere Formen der Zustimmung wie *sicher*, *natürlich*, *richtig*.
- Zu den Merkmalen der Verwendung eines zweiteiligen und in sich oppositiven Formativs gehört, dass der erste Teil einen zweiten, adversativen Teil projiziert. Das geschieht mit Verfahren der Zustimmungseinschränkung und Relevanzrückstufung. Die Projektion des oppositiven zweiten Teils ist in vielen Fällen sehr deutlich; sie gehört zum Musterwissen kompetenter Interaktionsteilnehmer und wird z. B. auch in kommunikativen Spielen verwendet.⁵

Da *ja* für den ersten und *aber* für den zweiten Teil mit Abstand die häufigsten Markierungen sind, wird als allgemeine Kennzeichnung der Funktionen der beiden Teile JA und ABER verwendet, unbeschadet der Tatsache, dass es Funktionsunterschiede in Abhängigkeit von den gewählten Konnektoren und sonstigen Formulierungsmustern gibt.

Der erste Teil des Komplexes kann unterschiedlich ausgebaut sein. Neben minimalen Fällen wie *ja aber eh *kaprizieren sie sich nicht auf die neue Prüfungsordnung* gibt es auch Fälle, in denen die Ausdrücke für JA und ABER

„Initiative Reaktionen“ im SFB 245 „Sprechen und Situation“. Vgl. dazu die beiden SFB-Papiere Kallmeyer/Schmitt 1991a und b. Diese frühen Fassungen werden zur Zeit in grundlegend überarbeiteter Form für die Veröffentlichung vorbereitet.

⁵ In einem Comic aus der Serie „Hägar der Schreckliche“ tritt Hägar mit seinem Adlatus einer größeren Gruppe von sehr angriffslustig wirkenden Feinden entgegen und begrüßt sie: Ah ... das **Empfangskomitee!** – Okay! Ihr seid uns vielleicht **zahlenmäßig** überlegen ... vielleicht seid ihr auch **größer!** – Vielleicht seid ihr **stärker** und **zäher** als wir! – Vielleicht könnt ihr uns sogar **besiegen!** – An dieser Stelle schaltet sich der Adlatus ein: *Kommst du bald zu einem „Aber“?*, worauf Hägar irritiert und ängstlich zur Seite blickt und entgegnet: *Ich dachte, das wäre ich schon!*

einen großen Wortabstand haben. Im folgenden Beispiel beträgt er 30, aber der im Korpus festgestellte Höchstwert ist 59.⁶

- 86 RS: weiß jetzt nich ob es ne vorlesung is jetzt oder
 87 BR: |s=is en proseminar- das is en proseminar| was
 88 RS: |n seminar ↓ oder n seminar ↓ |
 89 BR: rodenberger macht ↓
 90 RS: also sie würden eher an=ne ne
 91 BR: <äh **na ja** RÄUSPERT SICH |im
 92 RS: vorlesung- <ach so |ja
 93 BR: prinzip is es so daß die | daß |die äh |
 94 RS: dann hab doch hab ich auch noch>| |LACHT ETWAS|
 95 BR: daß man |einführung>| * äh * → und proseminar nicht im
 96 RS: |> (... ...) <|
 97 BR: gleichen semester machen soll ↓ sondern erst die
 98 BR: einführung dann das proseminar ↓ ←
 99 RS: ah ja ↓ * ja dann lass
 100 BR: |**aber** | äh was rodenberger über kafka
 101 RS: ich das fa|llen ↓|
 102 BR: macht- * dürfte auch einführenden charakter haben **

Das Beispiel zeigt u. a., dass JA-ABER-Konstruktionen in eine Sequenz eingebettet sind, zu der eine Anlass- bzw. Bezugsäußerung gehört (vgl. *also sie würden eher an=ne vorlesung-*, 90/92) und oft auch Folgeschritte der beteiligten Sprecher.

Die erwähnten Annahmen über die Mustereigenschaften sollen anhand einer größeren Materialmenge weiter belegt, ggf. weiter ausdifferenziert und auf ihre Nutzbarkeit für eine automatische Recherche hin untersucht werden. Um das Vorgehen plausibel und fruchtbar zu machen, soll die Materialmenge so groß sein, dass sie nicht ohne weiteres mehr im Kopf präsent gehalten werden kann. Zugleich muss das Material zumindest in größeren Teilen auch von Hand analysiert werden. Dazu empfiehlt sich ein Auswahlkorpus von wiederum begrenztem Umfang. Das hier benutzte Korpus enthält insgesamt 87.629 laufende Wörter. Die Erfahrung zeigt, dass dieser Umfang beide Bedingungen erfüllt und eine praktikable Größe als exploratives Versuchskorpus darstellt. (Vgl. die Korpusübersicht im Anhang).

⁶ Das Gesprächskorpus des IDS benutzt ein Partitursystem mit fortlaufenden Sprecherzeilen, ohne interpretative Segmentierungen (z. B. in Äußerungen). Die Transkriptionen haben einen mittleren Feinheitsgrad. Bei Überlappungen von Redebeiträgen und prosodischen Markierungen von Lautstärke und Tempo werden jeweils Anfang und Ende notiert; das hat u. a. Vorteile für die Recherchierbarkeit. (Vgl. die Transkriptionserläuterung im Anhang).

Das Korpus enthält 10 Beratungsgespräche unterschiedlicher Art (Studienberatungen, ärztliche Beratung, Finanzberatung, Sozialberatung), eine Referatsbesprechung zwischen Professor und Student, ein universitäres Prüfungsgespräch, 3 Fernsehdiskussionen und eine Arbeitsbesprechung. Die Zusammensetzung des Korpus mit einem relativ großen Kernbereich (Beratung) und einigen näher oder ferner liegenden Kontrastfällen entspricht einem explorativen Vorgehen zur Hypothesenentwicklung in qualitativen Untersuchungen. So wurde das vorliegende Korpus zusammengestellt, um Hypothesen über die Rolle von adversativen Konstruktionen in helfenden Interaktionen als Folge der Beratungsaufgabe zu entwickeln. Zu Beratern gehört die Nutzung von unterschiedlichen Perspektiven als Lösungsresource und damit die Aushandlung der Perspektivenunterschiede bzw. auch die Durchsetzung einer Perspektive. Insofern liegen immer auch Verfahren der Durchsetzung „zum guten Zweck“ nahe. Eine gesprächsanalytisch und auch für die Beratungspraxis interessante Frage ist, wie mit dem Durchsetzungspotential so umgegangen wird, dass es für Helfen funktional ist und nicht „Recht behalten“ dominant wird. (Vgl. u. a. Kallmeyer 2000).

Die folgende Betrachtung konzentriert sich auf den Zusammenhang von JA + ABER mit den oben skizzierten Formativ-Eigenschaften. Auf einige Erscheinungsformen, die Ähnlichkeiten haben, aber ggf. dem Zielprofil nicht ganz entsprechen, soll hier noch hingewiesen werden. Sie werden bei den ersten Recherche-Schritten mit erfasst, aber anschließend ausgeklammert. Die strukturellen Verwandtschaften und Unterschiede dieser Erscheinungsformen im Verhältnis zu den JA-ABER-Fällen sind gesondert zu untersuchen. Das sind insbesondere folgende Vorkommen von ABER-Konnektoren:

(a) ABER als Diskursgliederung

- 155 BR: also das-| ** aber äh * möglichst >wenig ↓ <
 156 RS: | gut ↓ **
 157 BR: | >gut ≤ |
 158 RS: **aber** ich glaub das wär=s dann |nich| ** ham sie

Der Ausdruck *gut* fungiert hier als Erledigungs- und Abschlussmarkierung, und die Äußerung mit *aber* formuliert die Erledigungsfeststellung noch einmal explizit.

(b) Initiales ABER ohne erkennbare Einräumung

Unter den 564 ABER-Vorkommen im Korpus finden sich 29 Kandidaten für initiales ABER. Diese Fälle sind interessant, weil sie in der Regel in Kontexten von andauernden Kontroversen erscheinen und insofern ein Dissensmerkmal darstellen. In dieser Hinsicht ist das entscheidende Kriterium für die Klassifikation als initiales ABER, dass die vorausgehende Äußerung desselben Sprechers nicht als JA-Teil einzustufen ist. Weiter zurück wird man in den meisten Fällen bei der Etablierung des Dissenses auch JA-ABER-Kon-

struktionen finden. Das ist auch beim folgenden Ausschnitt aus einer psychologischen Beratung der Fall:

- 180 BL: paniksätzen fütterst ** die hätte ich ganz gerne
 181 BL: >diese paniksätze-< *2,5*
 182 MA: ja eigentlich sind sie
 183 BL: |wo du| ni"cht schnaufen
 184 MA: aggressiv diese paniksätze |also |
 185 BL: kannst *2*
 186 MA: ja also sie lösen ja dann eine art wu"t
 187 BL: **aber** bei der wut ka"nnste doch
 188 MA: dann in mir aus
 189 BL: wieder schnaufen ** ich möchte gerne die"
 190 MA: ja↑ **
 191 BL: sätze hörn wo du ni"cht schnaufen kannst↓

Es gibt auch spezifische Verwendungen von ABER ohne Vorgänger-JA als Instrument der Gesprächssteuerung, z. B. bei Moderatoren von Diskussionen wie Sabine Christiansen. In der Fernsehdiskussion mit dem Kandidaten Stoiber formuliert sie die meisten ihrer Fragen ebenso wie thematisch steuernde Kommentare mit initialem ABER (insgesamt 9 mal).

(c) Es gibt eine gewisse Anzahl von Fällen, in denen ein Konnektor vom Typ ABER erscheint, aber nicht erkennbar ist, dass der vorausgehende Äußerungsteil eine Projektion eines oppositiven zweiten Teils enthält. Vielmehr wird in diesen Fällen in einer Äußerungserweiterung retrospektiv eine Oppositionsrelation eingeführt. Im folgenden Beispiel aus dem universitären Prüfungsgespräch hat der Student (BB) in einem langen Redebeitrag eine Textstelle interpretiert. Am Ende dieser Darstellung schließt er eine Äußerungserweiterung mit *allerdings* an, die in der vorausgehenden Äußerung durch nichts erwartbar gemacht wurde:

- 141 BB: ** und u:nd ** und hier wird eben der
 142 BB: zweikampf alexander menes geschildert- *
 143 AA: ja
 144 BB: bis achtundfünfzig- be/ bis
 145 BB: achtundfünfzig * den der alexander
 146 BB: **allerdings** verliert↓ ** den schwertkampf
 147 AA: aha↑ ** ←ist das→ typisch für
 148 BB: verliert er↓
 149 AA: das alexanderbild↑

Interessanterweise wird dieser „unmotivierte“, d. h. nicht projizierte oppositive Zusatz Gegenstand einer interaktiven Bearbeitung. Dabei spielen zwei Aspekte eine Rolle. Zum einen fällt auf, dass die Einschränkung bzw. Relativierung in einem abhängigen Satz (Relativsatz) erscheint. Ohne sonstige Hochstufungen, z. B. prosodischer Art oder in Form einer Fokussierungsformel, ist das für JA-ABER wichtige Merkmal der Relevanzhochstufung im zweiten Teil nicht gegeben. Immerhin bearbeitet BB diesen Aspekt nachträglich dadurch, dass er den Einschränkungssachverhalt noch einmal als Hauptsatz und damit als eigenständigen Sachverhalt formuliert. Dieses Verfahren ist vermutlich so zu interpretieren, dass der Sprecher ein thematisches Potential einführt und mit der Reformulierung in der Relevanz hochstuft. Wie die Reaktion von AA zeigt, sieht dieser Anlass, genau diesen Punkt mit seiner Nachfrage zu bearbeiten. Einerseits erfolgt durch *aha*↑ eine markierte Aufmerksamkeitszuwendung, also eine Relevanzhochstufung, und andererseits wird ein möglicherweise in Opposition stehender Sachverhaltsaspekt explizit eingeführt.

(d) Die Kombination von Zurückweisung + ABER, realisiert als *nein aber* mit der Funktion einer Voraussetzungskorrektur. Sie kommt im Korpus mehrfach vor (vgl. die Korpusübersicht im Anhang). Im folgenden Beispiel hat eine Studentin die Lösungsvorstellung, dass sie nach einer noch in Vorbereitung befindlichen neuen Prüfungsordnung Examen machen könnte, weil diese Prüfungsordnung für sie möglicherweise günstiger ist. BR erläutert die Unklarheit der Situation (*das is dann noch nich geläufig jetzt weil wir die neue prüfungsordnung noch nich haben*) und stellt die Nachteile der Lösung „neue Prüfungsordnung“ dar. Die Reaktion von RS enthält zwei Aussagen: *das macht ja nichts* und *das wollt ich jetzt grad wissen was da*. BR wiederum reagiert mit einer Zurückweisung bzw. Korrektur von Voraussetzungen, die RS in ihrer zweiten Aussage benutzt (die Möglichkeit, über die neue Prüfungsordnung Auskunft zu geben), eingeleitet mit *nein* und einer Richtigstellung der Voraussetzung:

- 39 BR: kö"nnen sie zwar theoretisch optieren für die neue
 40 BR: prüfungsordnung↑ aber die wird so viel andere
 41 RS: >mhm<
 42 BR: anforderungen enthalten↓ |die
 43 RS: ja das macht ja nichts |das
 44 BR: sie-| |ja |
 45 RS: | LACHT wollt ich jetzt grad wissen |was| da was
 46 BR: **nein** →das kann ich ihnen natürlich auch nicht
 47 RS: da-
 48 BR: sagen weil ich die neue prüfungsordnung nicht
 49 BR: kenne↓← *|aber | es sind zum beispiel dann also
 50 RS: |ah so↑|

- 51 BR: vier hauptseminarscheine erforderlich-
 52 RS: in der neuen
 53 BR: |insge|samt * statt bisher zwei"
 54 RS: abteilung |(...)|

Mit der Richtungsstellung der Voraussetzung gibt BR seine Handlungsbedingungen an, im konkreten Fall seine mangelnden Voraussetzungen für eine vollgültige Antwort. Mit dem ABER-Teil unternimmt er aber dennoch einen Antwortversuch. Dabei reagiert er auf die Annahme von RS, dass die neuen Anforderungen mit ihren Lösungsvorstellungen kompatibel sein könnten. Die Behandlung eigener Voraussetzungen in Einräumungsteilen kommt häufiger vor, und solche Fälle werden auch berücksichtigt. Nur die Fälle mit Voraussetzungskorrekturen oder -zurückweisungen mit der Negationspartikel *nein* sollen hier vorerst ausgeklammert werden.

3. Wortsuche nach Konnektoren

3.1 Der ABER-Teil

In einem ersten Schritt wurden die Konnektoren für ABER im Korpus gesucht. Ihre Anzahl ist begrenzt.

Im vorliegenden Fall wurden in die Recherche sowohl adversative als auch konzessive Konnektoren aufgenommen, die sich dadurch auszeichnen, dass sie eine Relation der Inkompatibilität zwischen den zwei verknüpften Aussagen herstellen.⁷ Dabei kommen Konjunktionen wie *aber* und *obwohl* und auch Adverbien wie *nur*, *bloß*, *allerdings* in Betracht. Außerdem werden auch Konnektoren wie *auf der einen / anderen seite* bzw. *zum einen / anderen* berücksichtigt, welche die Unterschiedlichkeit von zwei verknüpften Aussagen markieren, weil sie tendenziell mit einer Relevanzhochstufung des zweiten Teils verbunden sind und diese häufig auch noch durch *aber* oder einen anderen Konnektor ausgedrückt wird; oft wird auch nur ein Teil des paarigen Konnektors verwendet. Für alle Ausdrücke wurden Aussprachevarianten berücksichtigt (wie *abba aba awwa awa* für *aber* oder *andre* für *andere* usw.) sowie alle Formen der in den Transkriptionen benutzen prosodischen Markierungen für Akzentuierung, Dehnung, größere Lautstärke und langsames Tempo.⁸ Die Ausdrücke *eine / andere seite* in den möglichen Varianten *die andere seite* bzw. *auf der anderen seite* sind mit *ander** bzw. *andre* seite* mit Wortabstand 1 problemlos recherchierbar.⁹

Auch *doch*, *jedoch* wurden am Material geprüft. Allerdings gab es keine Funde, in denen in einer zweiteiligen Äußerungskonstruktion die Opposition

⁷ Vgl. Zifonun/Hoffmann/Strecker et al. (1997), Kapitel H2.

⁸ Also: *a"ber* und *a:ber*, *<aber [xxx]>*, *←aber [xxx]→*; [xxx] bezieht sich hier darauf, dass die zweite korrespondierende Markierung sowohl direkt nach dem Konnektor als auch später erscheinen kann.

⁹ Vgl. die COSMAS-Operatoren *WORT(x)*, *BETONUNG*, *DEHNUNG*, *GLEICH(X, Y)*.

2295 VF: |auf die fra"ge↑| |LACHANSATZ|
 2295 AS: |mu"ss sie drauf| au"fmerksam |machen | ** es

Syntaktisch integriertes *aber*, *nur*, *bloß* oder *allerdings* erscheint innerhalb des ersten Syntagmas.¹¹ Um auch in diesen Fällen die unmittelbare Nähe des Konnektors zur Umschlagstelle zum ABER-Teil feststellen zu können, muss man Hilfsstrategien mit einbeziehen. Die Beobachtungen am Korpus haben ergeben, dass die erste Einheit von ABER mit integriertem Konnektor fast durchweg das Format einer einfachen und kurzen Konstruktion hat:

299 BR: a"ndre frage wo sie ihren lati"numschein machen
 300 BR: * das is völich ihnen überlassen↓
 301 RS: **mhm** *
 302 BR: |also ich mein ob sie dann das| mü"ssen sie nicht in
 303 RS: |**ja** das geh"t **aber** schwe:r- |
 304 BR: dreibug machen
 305 RS: **ja** das is **aber** dann nun ganz
 306 RS: schwierig wenn ich hier studiere und dann * ich kann

Hier verwendet RS ein Konstruktionsmuster, das für einen unmittelbar bezogenen Kommentar charakteristisch ist mit einer Proform vom Typ *das* bzw. *da* und Kompositionen damit wie *dabei*, *darum* usw. Dieses Formulierungsmuster ist kurz und in besonderer Weise dafür geeignet, den syntaktisch integrierten Konnektor nah an der Umschlagstelle zu platzieren.

Vergleichbare Beobachtungen kann man auch an anderen Beispielen machen, in denen die Markierung des JA-Teils weit entfernt und manchmal auch undeutlich ist:

209 BR: **ja** * **ja** * es is also keine *
 210 RS: alles eigentlich↓
 211 BR: schwierigkeit s=is **nur** so wenn sie jetzt schon im
 212 BR: sechsten semester sind also * ich meine es is jetzt

Hier wird zunächst mit *ja* * *ja* und einer negativen Zustimmung/Einräumung Teil 1 gebildet. Der ABER-Teil beginnt dann mit einer Fokussierungsformel mit eingebettetem *nur*. Vergleichbar wären *es ist aber so* oder *es ist bloß so*. Auch dieses Formulierungsformat stellt die Nähe des Konnektors zur Umschlagstelle her. Im Unterschied zur Kommentar-Formel, die sich durch einen deutlichen und deiktisch-unmittelbaren Rückbezug auszeichnet, dominiert bei der Fokussierungsformel die kataphorische Funktion (in der Regel mit Hauptakzent auf *so* und Intonationsbruch). Die Suchvorgabe für solche Fälle

¹¹ Vgl. die Positionsklassen der Konnektoren in Pasch/Brauß/Breindl (2004), S. 502 ff.

Noch schwächer hinsichtlich der Bestätigungsfunktion als *ja* sind Rückmeldungen wie *hm* oder *mhm*. Eine praktikable Entscheidung scheint hier zu sein, dass einfache Rückmelder ohne besondere Markierung nur als Kandidaten einbezogen werden, wenn sie in kurzem Abstand dem ABER-Teil voraufgehen. Dieser Abstand könnte z.B. auf vier Wörter Diskurs (d. h. von einem anderen, gerade aktiven Sprecher) festgelegt werden. Ausscheiden könnte man als leise gesprochen transkribierte Rückmeldungen (z. B. >*mh*<).

Bei anderen Markierungen des ersten Teils durch markante und ggf. markierte Formen (also nicht durch nicht hervorgehobenes *ja*) ist ein Wortabstand von 10–15 in der Regel für die Recherche kein Problem, d. h. die zurückgelieferten Fälle sind mit großer Wahrscheinlichkeit echte Treffer. In dieser Hinsicht ist das folgende Beispiel ganz eindeutig:

- 521 BR: semester bequem machen↓
 522 RS: **eben**↓ und deshalb da fehlt"
 523 BR: |ja |
 524 RS: mir ja nich mehr so |viel| * **obwohl** in traun
 525 BR: |RÄUSPERT SICH |
 526 RS: |fehlt mir dann | wieder mehr" die haben *

Der Konnektor *obwohl* erscheint hier in syntaktisch herausgelöster Frontstellung. Das ist für das Recherchesystem allerdings nicht erkennbar. Aber die Zusammengehörigkeit von *eben*↓ am Äußerungsbeginn und *obwohl* im Wortabstand 11 ist so wahrscheinlich, dass eine solche Stelle problemlos als Treffer behandelt werden kann.

Als Leithypothese kann man formulieren: Je größer der Abstand zwischen JA und ABER, desto größer ist der Bedarf an deutlichen Markierungen. Das bedeutet für die Recherchemöglichkeiten: Markierte Formen erlauben einen größeren Wortabstand. Im folgenden Beispiel aus „Luxemburgerin“ wird z. B. der Zusammenhang von JA und ABER durch Mehrfachmarkierung von JA gestützt:

- 480 BR: |würden sie | ein:
 481 RS: gehört hab das is ganz |deprimierend-|
 482 BR: deutsches staatsexamen dann auch anerkannt bekommen
 483 BR: in luxemburg↓ * |für also- gymnasialbereich↓|
 484 RS: |**ja** das" **so**"**wieso**- das | is
 485 RS: also das **auf jeden fall: ja** das wird" anerkannt **bloß**
 486 RS: die lu"xemburger examen werden ja in deutsch"land
 487 BR: ja: ja |ja| ja
 488 RS: nich anerkannt↓ |es| wird schon anerkannt

Hier ist der JA-Teil vielfältig durch Zustimmungsausdrücke besetzt: *ja, so" wieso, auf jeden fall, ja*. Der Wortabstand ist relativ gering, aber wenn man Fälle mit Mehrfachmarkierung des ersten Teils sucht, bekommt man auch bei großem Wortabstand echte Treffer.

Ein typischer Treffer für Mehrfachmarkierung im JA-Teil ist auch das folgende Beispiel („Luxemburgerin“):

306 RS: schwierig wenn ich hier studiere und dann * ich kann

307 RS: ja nicht da und hier sein LACHT ich kann ja nich

308 RS: dauernd * mich hier * hin und her fahren

309 BR: |**natürlich** also äh | für des: für=s

310 RS: |mit=m auto des geht ja kaum↓|

311 BR: ler"nen für die * äh für=s latinum * **das is klar-**

312 BR: **nur** ich meine wo sie die abschlussklausur

313 BR: schreiben↓

Die beiden Ausdrücke *natürlich* und *das is klar* haben einen Wortabstand 12 und werden mit erfasst, wenn man ausgehend von *nur* mit einem Wortabstand von 15 (einschließlich der Pausenfüller *äh*) Manifestationen von JA sucht. Der Ausdruck *das is klar* wird durch die Wortsuche nach *klar* erfasst, kann darüber hinaus aber ggf. auch als Bestandteil eines Kommentarmusters identifiziert werden. Der Fall zeigt zudem eine interessante Eigenschaft von Einräumungsteilen: Das erste Adverb der Zustimmung (*natürlich*) erscheint am Äußerungsbeginn und *das ist klar* unmittelbar vor dem ABER-Konnektor. Dies entspricht einer Rahmung des JA-Teils. Solche Rahmungen können maschinell gefunden werden, wenn der Beginn der Einräumung identifizierbar ist wie im vorliegenden Fall (Äußerungsbeginn). Beginnt die Einräumung irgendwo im Verlauf einer längeren Äußerung, können Segmentierungsmuster hilfreich sein, aber die müssen erst noch systematisch erarbeitet werden.

Im folgenden, bereits eingangs zitierten Beispiel mit dem Wortabstand 30 zwischen *ja* und *aber* gibt es ebenfalls zusätzliche stützende Elemente im JA-Teil:

86 RS: weiß jetzt nich ob es ne vorlesung is jetzt oder

87 BR: |s=is en proseminar- das is en proseminar| was

88 RS: |n seminar↓ oder n seminar↓ |

89 BR: rodenberger macht↓

90 RS: also sie würden eher an=ne ne

91 BR: <äh na ja RÄUSPERT SICH |im

92 RS: vorlesung- <ach so |ja

- 93 BR: **prinzip is es so dass** die | dass |die äh |
 94 RS: dann hab doch hab ich auch noch>| |LACHT ETWAS|
 95 BR: dass man |einführung>| * äh * → und proseminar nicht im
 96 RS: |>(… …)< |
 97 BR: gleichen semester machen soll ↓ sondern erst die
 98 BR: einführung dann das proseminar ↓ ←
 99 RS: ah ja ↓ * ja dann lass
 100 BR: |**aber** | äh was rodenberger über kafka
 101 RS: ich das fall|len ↓
 102 BR: macht- * dürfte auch einführenden charakter haben **

Eine Stütze für die Erkennbarkeit des Zusammenhanges von BRs Äußerung ab 91 und seiner vorausgehenden Äußerung ist, dass BR den Beginn seiner Äußerung durch die expandierte *ja*-Form *na ja* markiert, der man aufgrund des *na* zudem eine Relativierungsprojektion zuschreiben kann; weiter durch den möglichen Oppositionsanker *im prinzip* sowie eine Ankündigung bzw. Fokussierungsformel *is es so dass*. Aus einer gesprächsanalytischen Perspektive würde man zudem die besondere Gestaltung des Beginns als ein Hervorhebungsverfahren mit projektiver Funktion ansehen: Anhebung der Lautstärke, Verzögerung *äh* + *na ja* + weitere Verzögerung durch Räuspern. Ob ein solches Markierungsmuster für die Recherche weitere Fälle erschließt, die sonst zweifelhaft geblieben wären, kann allerdings nur anhand eines größeren Korpus geprüft werden.

Weiter gibt es eine Reihe von Fällen, in denen die Einräumung unter Verwendung von Verben des Wahrnehmens, Denkens und Sagens formuliert wird. In diesen Zusammenhängen wird die Übereinstimmung häufig durch *auch* und Äquivalente markiert: *das denke ich auch*; *das glaub ich au"ch* ↓ * *ähm aber* (Sklavenreferat 406/407) und ähnliche Fälle.

Spätestens an dieser Stelle wird jedoch deutlich, dass die Recherche von relativ variablen Formulierungsmustern an ihre Grenzen stößt. Die Ausdrücke für Einräumung sind zahlreich und strukturell unterschiedlich. Neben den bisher genannten kommen noch viele in Betracht, die Wahrnehmung, Denken, Fühlen und Handeln des Sprechers betreffen, z. B. *das sehe ich auch so* usw. Unter den gegebenen Recherchebedingungen verspricht eher Erfolg, solche Einräumungsmuster ausgehend von bereits über Konnektoren und andere geläufige Ausdrücke identifizierte JA-ABER-Fällen zu eruieren. Für die erwähnten Fokussierungsformeln ist die Lage insofern besser, als einige von ihnen gut auffindbar sind.

Es gibt weitere Fokussierungsformeln, die für die Relevanzhochstufung an der Schnittstelle zwischen JA und ABER verwendet werden. Kandidaten sind *es geht um*, Formulierungen mit *folgendes* und auch Formulierungen mit *dazu* / *hier* / *an dieser stelle* + flektierte Form eines Modalverbs (*müssen* / *wollen* /

können / sollen, vor allem *sollte, sollten*) + Subjekt *ich / wir / man*, ggf. auch andere personale Referenz in der dritten Person + Verbum dicendi im Infinitiv (*sagen / erwähnen* usw.). So setzt im Beispiel auf S. 212 AS den ABER-Teil fort mit *ich mu"ss sie drauf au"ffmerksam machen*. Neben den positiv formulierten Konstruktionen gibt es auch negative Äquivalente wie *wir sollten / man darf nicht verschweigen* usw.

Die Menge der Formulierungsmuster für Einräumung ist relativ groß, und dementsprechend selten kommen die einzelnen Muster vor. Um sie zu eruieren, braucht man ein wesentlich größeres Korpus als das hier benutzte. Wenn man mit großen Listen von möglichen Mustern arbeitet, wird das Recherche-Verfahren zudem schwerfällig. Der Aufwand lohnt sich allerdings bei sehr großen Korpora.

Ähnlich wie die Fokussierungsformeln können auch bestimmte Segmentierungsausdrücke für die Recherche hilfreich sein. In vielen Fällen steht die Einräumung nicht am Äußerungsbeginn. Wenn sie auch keine der gut recherchierbaren Ausdrücke für JA enthält, bleibt nur die Recherche von Segmentierungsformeln. Im folgenden Beispiel („Luxemburgerin“) beginnt der Sprecher zunächst eine ABER-Formulierung, bricht sie ab, platziert eine Einräumung und setzt dann mit ABER fort. Der Sprecher benutzt *ich meine* als typische Segmentierungsformel:

211 BR: schwierigkeit **s=is nur so** wenn sie jetzt schon im

212 BR: sechsten semester sind also * **ich meine** es is jetzt

214 BR: nich im nächsten semester notwendig **aber** so langsam *

216 BR: müßten sie sich mal * |um | ihren prüfer *

217 RS: |latein|

Vgl.

218 BR: kümmern * neben dem latein ↓

219 K LACHEND

Der Ausdruck *ich mein** (d.h. mit Deklinationsmorphemen oder ohne) ist gut recherchierbar. Er erscheint in den allermeisten Fällen als Eröffnungsmarkierung von Parenthesen und Expansionen (manchmal allerdings auch bei der Bearbeitung von Verständigungsproblemen, z. B. *nein ich meine*), und er erscheint relativ häufig am Beginn von Einräumungen (z. B. *ich mein ich weiß auch nicht wie=s im au"genblick aussieht nur also ich würd anrufen...*, „Wiedereingliederung“ 297; *nich das ich mein d=s natürlich glücklich dass es so wenige sind ds macht viel aus aber wir machen es uns ...*, „Sklavenreferat“ 1223-8). Im gesamten hier benutzten Korpus ist unter 25 Fällen von *ich mein** im Wortabstand bis 15 vor ABER nur eine Verständigungskorrektur (*nein ich meine ... aber*, „Wiedereingliederung“ 217), während in allen anderen Fällen *ich meine* den Einräumungsbeginn markiert. Insofern kann die Recherche von

Formulierungselemente bei einer formalen Analyse Zusammengehörigkeit signalisieren kann (*zum ei"nen is es* und *und da"nn is es*). Beim gegenwärtigen Kenntnisstand ist keine klare Entscheidung möglich. Es wird notwendig sein, anhand wesentlich größerer Korpora zu überprüfen, ob und ggf. mit welcher Wahrscheinlichkeit *dann* als zweiter Konnektor zu *zum einen* und ggf. anderen JA-Konnektoren vorkommt. Kombiniert mit *aber* ist das häufiger der Fall (*aber dann*), aber dabei gibt es auch für die Suche des ABER-Konnektors kein Problem.

Trotz aller Erfolge der bisher entworfenen kombinierten Recherche von Konnektoren und häufigen JA-Markierungen unter Zuhilfenahme von zusätzlichen Markierungen und Formulierungsmustern ist die Zahl der nicht einfachen Fälle im Korpus immer noch beträchtlich. So bietet auch der folgende Fall einige typische Schwierigkeiten:

1575 SC: |können |s|ie: * ja" oder nei"n sagen dass sie

1576 ES: da|rüber muss|

1577 SC: sagen * ich möchte es gerne aus=em wahlkampf

1578 SC: rau"shalten * ich möchte eine schne"lle einigung *

1579 SC: oder * wird=s innen wahlkampf gehen↓ **

[ES antwortet nicht direkt auf die Frage]

1585 SC: |also| ja" sie

1586 ES: wi"rklich nicht das wird immer wie|der |

1587 SC: möchten sich gerne * |zü"ig einigen |

1588 ES: |ja wenn wenn die| s p d *

1590 ES: auf unsere positionen * die wir schon früh" im

1591 ES: mai" des letzten jahres- ** c d u c s u↓ frau

1593 ES: merkel und ich haben das vorgestellt * die

1594 ES: eckpunkte von c d u c s/ wir waren ja früher dran

1596 ES: als herr schily- * wenn diese dinge * eingehalten

1597 ES: werden↑ * äh übernommen werden↑ * äh da"nn↑ äh: *

1598 ES: **seh ich überhaupt kein proble"m * nu"r-** * ich

1599 ES: möchte nicht irgendwelche au"sführungen von herrn

1600 ES: schily haben↑ * sondern ich möchte den

1601 ES: gese"tzestext haben↑ * ich möchte wi"ssen↑ was *

In diesem Fall ist der ABER-Konnektor markiert (Pause + Akzentuierung + Pause) und wird insofern bei der Wortsuche sofort erkannt. Im unmittelbaren

Kontext von *nu"r* erscheint die Formulierung *seh ich überhaupt kein proble"m*. Hierin steckt als Formulierungsmuster *kein problem*. Der Ausdruck *problem* alleine kommt vielfältig vor, teilweise auch als Fokussierungselement (*das proble"m ist*), aber auch in vielen anderen Kontexten. Der kombinierte Ausdruck *kein problem* ist seltener und wird spezifischer verwendet. Es gibt im Korpus eine Reihe von Fällen, in denen *kein problem* im Zusammenhang mit Oppositionsrelationen erscheint und dann immer im ersten Teil steht: *hab ich kei"n problem * ähm ** a:ber *3* >ä:hm< *4* für mi"ch ...* („Feldbusch-Schwarzer“ 212); *ich wollte ihnen kein problem ... einreden ich wi"ll nur sagen ...* („Feldbusch-Schwarzer“ 1369–74). Insofern kann *kein problem* unter die Formulierungsmuster für Einräumung aufgenommen werden. Im vorliegenden Fall wird die Stelle bei einer Suche nach ABER+ JA-Markierungen (anhand einer Liste relevanter Formen) gefunden.

Damit ist aber noch nicht der Beginn der Einräumung erkannt. Das Problem ist hier, dass das initiale *ja* unmarkiert und prosodisch integriert ist. Es kann als eine Markierung für Eröffnung und Bezugnahme angesehen werden, aber kaum als Zustimmungsausdruck. Hinzukommt der große Wortabstand von 59, wenn man die einzelnen Elemente der Abkürzungen für die politischen Parteien als Wörter zählt (so wie sie transkribiert sind). Interpretativ spielt bei der Analyse für das frühzeitige Erkennen einer Einräumung natürlich der vorausgehende Kontext eine große Rolle. SC hat Zustimmung manifest relevant gesetzt, indem sie die vorausgehenden Äußerungen von ES in Reaktion auf die frühere SC-Frage als *also ja" sie möchten sich gerne * zügig einigen* reformuliert. In diesem Kontext ist auch das unmarkierte *ja* von ES als Behandlung der Akzeptanz bedeutsam. Man erkennt sofort, dass hier mit *ja wenn* eine bedingte Zustimmung formuliert wird, die in den schon erkannten Einräumungsteil mündet: ** äh da"nn↑ äh: * seh ich überhaupt kein proble"m*. Das weiß aber das Recherchesystem nicht.

Eine auf diesen Fall zugeschnittene Suchstrategie könnte sein, ausgehend von der Einräumungsformel im unmittelbar vorausgehenden Kontext Konnektoren zu suchen, die das Potenzial haben, weitgespannte Formulierungszusammenhänge zu verklammern, sowie ggf. mögliche Partner für paarweise verwendete Konnektoren. Dabei stößt man dann auch auf *wenn-dann*. Wenn sich eine Anbindung über einen solchen Konnektor findet, könnte man die Kontextspanne bis zum Äußerungsbeginn öffnen. Um daraus ein praktikables Suchverfahren zu entwickeln, braucht man allerdings wiederum ein wesentlich größeres Korpus. Immerhin zeigen die bisherigen Erfahrungen mit der Recherche von Formulierungsmustern, dass solche Hilfsstrategien durchaus erfolgreich sein können.

4.3 Relative Häufigkeiten von Wörtern in begrenzten Kontexten

Um die bisherigen Möglichkeiten der maschinellen Recherche von komplexen Zusammenhängen für JA-ABER-Sequenzen einschließlich Anlassäußerung und

möglichen Anschlussäußerungen weiter abzusichern, sollte man auch andersartige Verfahren in Betracht ziehen. Die bisherigen Recherche-Verfahren gehen von festgelegten Wörtern bzw. Wortformen aus. Nun kann es auch hilfreich sein, nach Verteilungsmustern von vorher nicht bestimmten Wörtern zu suchen.

An einer Reihe von Gesprächsausschnitten mit JA-ABER-Konstruktionen fällt auf, dass in dem begrenzten Kontext dieser Konstruktion einzelne Ausdrücke mehrfach wiederholt werden, die im Gespräch insgesamt selten sind und ggf. nur in diesem lokalen Kontext vorkommen. Diese Wortwiederholungen sind Teil der engen inhaltlichen Bezugnahme zwischen Anlassäußerung und JA-ABER-Komplex. Besonders deutlich wird das in den Fällen, in denen alle Aktivitätsschritte in einer Interaktionssequenz jeweils das entsprechende Wort enthalten. Im folgenden Beispiel („Luxemburgerin“) gibt es zwei Kandidaten für eine solche Verdichtung in der Wortverwendung, *examen* und *anerkannt*, wobei *anerkannt* mit vier Vorkommen besonders hervortritt:

- 480 BR: |würden sie | ein:
 481 RS: gehört hab das is ganz |deprimierend-|
 482 BR: deutsches staat**examen** dann auch **anerkannt** bekommen
 483 BR: in luxemburg ↓ * |für also- gymnasialbereich ↓|
 484 RS: |ja das" so"wieso- das | is
 485 RS: also das auf jeden fall: ja das wird" **anerkannt** bloß
 486 RS: die lu"xemburger **examen** werden ja in deutsch"land
 487 BR: ja: ja |ja| ja
 488 RS: nich **anerkannt** ↓ |es| wird schon **anerkannt**
 489 RS: aber es is eben bei uns auch so wie hier in
 490 BR: |ja | |ja:
 491 RS: deutschland un|heimlich|cher lehrerüberfluß |jetzt

Der Ausdruck *anerkannt* kommt in diesem Gespräch relativ häufig vor (und erweist sich insofern als Schlüsselwort für ein zentrales Thema der Beratung), insgesamt 11mal (bei insgesamt 5128 Wörtern, was 0.194% entspricht), aber nicht in der lokalen Verdichtung wie in der obigen Stelle. Es gibt nur noch eine Stelle mit zwei Vorkommen in zwei aufeinander folgenden Äußerungen („Luxemburgerin“ 828+829).

Die ersten drei Vorkommen erscheinen in den drei Schritten der Sequenz Anlassäußerung (Frage von BR) – Einräumung – ABER. Das vierte Vorkommen liegt in der Einräumung eines nächsten JA-ABER-Komplexes. In diesem Fall fungiert die vorausgehende ABER-Äußerung als Anlass bzw. Bezugsäußerung für eine weitere JA-ABER-Konstruktion. Die Wiederholung des Ausdrucks *anerkannt* markiert die inhaltliche Bezugnahme, aber – anders als im vorausgehenden JA-ABER-Komplex – nicht den internen Zusammenhang zwischen JA und ABER.

Der Ausdruck *examen*, als Simplex oder Teil eines Kompositums, kommt im Gespräch 13mal vor und an anderen Stellen auch in höherer Verdichtung. Im Prinzip ist ein zweimaliges Vorkommen die minimale Bedingung für die mögliche Relevanz eines solchen Ausdrucks als Schlüsselwort, allerdings nur in Kontexten, in denen das Wort sonst nicht oder nie verdichtet vorkommt.

Für solche Recherchen kann man die fortlaufenden Indices aller Ausdrücke in COSMAS benutzen und Abstände (z. B. entsprechend einem Kontext von 80 Wörtern) in Indexzahlen und Mindesthäufigkeiten (z. B. gleich/größer 3) vorgeben, die an der Gesamtmenge des Vorkommens relativ zur Wortmenge des Gesprächs gemessen werden (also z. B. alle Ausdrücke, die nicht mehr als 0,2 % des Gesamtvorkommens ausmachen). Für die hier durchgespielte Analyse von JA-ABER-Äußerungen wird man zudem diese Recherche auf die Umgebungen von ABER-Ausdrücken beschränken. Entscheiden muss man auch, ob man nur Wörter oder auch Wortteile berücksichtigen will. Eine Recherche der geschilderten Art würde z. B. zum Gespräch „Christiansen-Stoiber“ zurückliefern, dass im Gespräch der Ausdruck *kompetenzteam* insgesamt dreimal vorkommt und dass alle Vorkommen in einem begrenzten Kontext liegen:

- 334 SC: wenn ich
 335 ES: diensttag=äh: * den oder die vorstellen↓ *
- 336 SC: mir das **kompetenzteam** so a"nschae↑ dann denk ich
 337 SC: wo bleibt die kompetenz der frau"en↓ frau me"rkel
 338 SC: ist natürlich dabei- * aber das war=s dann auch
 339 SC: schon↓
 340 ES: ja das ist nicht ein **kompete"nzteam**↓ das
 341 ES: was wir jetzt=äh:: zunächst einmal festgelegt
 342 ES: haben↑ * äh das ist=äh * die führungsmannschaft
 343 ES: die den wahlkampf inhaltlich und organisatorisch *
 344 ES: äh * vorbereitet↑ * u:nd äh: **natürlich** ko"mmen
 345 ES: dann in einer späteren phase * kommt ein
 346 ES: **kompetenzteam**↑ * >äh:< mit einigen ausgewählten
 347 ES: persönlichkeiten die dann auch für einzelne *
 348 ES: sa"chfragen stehen↑ * **aber**- * das hat zeit
 349 ES: zunächst äh: geht es ja darum den wahlkampf zu
 350 ES: verzahnen↑ * bestimmte themen gemeinsam

Auch hier verteilen sich die Vorkommen des Schlüsselwortes auf drei Schritte der oppositiven Sequenz, die insgesamt aus einer Anlassäußerung von SC, einer Zurückweisung mit Richtigstellung und einer JA-ABER-Konstruktion von ES besteht. Der Ausdruck *kompetenzteam* erscheint in der Anlassäußerung (*wenn ich mir das kompetenzteam so anschau*), der Zurückweisung von ES (*ja das ist nicht ein kompetenzteam*↓ *das was wir jetzt=ä:h zunächst einmal festgelegt haben*) und seiner Einräumung (*u:nd äh: natürlich ko"mmen dann in einer späteren phase * kommt ein kompetenzteam*↑).

Wenn man bei der Recherche auch Wortbestandteile zulässt, bekommt man für „Christiansen-Stoiber“ u. a. folgendes Ergebnis: Für den Ausdruck *team* gibt es im Gespräch insgesamt 5 Vorkommen (was 0,041 % einer Wortmenge von 12.237 entspricht), die sich an zwei Stellen verdichten, die zudem dicht aufeinander folgen und zusammen eine Gesprächsstrecke von ca. 145 Wörtern ausmachen. ES verwendet in einer Äußerung zweimal den Ausdruck *team*, in den Verbindungen *teamkandidat* und *team stoiber*:

323 ES: wir haben uns da"zu entschieden dass wir * einen

324 ES: >äh< **team**kandidat ein **team** stoiber machen mit

325 ES: einem: * exzellenten * mann an der spitze der **

Auf diese Äußerung reagiert SC mit der im vorigen Ausschnitt erscheinenden Frage von SC, auf die wiederum ES mit seiner oppositiven Äußerung reagiert. Alle fünf Vorkommen von *team* stehen also in einem interaktiven Zusammenhang.¹²

Solche kleinräumigen Verdichtungen zeigen eine enge inhaltliche Bezugnahme und, im Zusammenhang mit JA-ABER-Konstruktionen, Aushandlungsprozesse. Solche Aushandlungszusammenhänge müssen nicht so knapp zusammengedrängt sein wie in den vorliegenden Fällen. Aber zumindest für bestimmte Aushandlungsmuster ist charakteristisch, dass die Beteiligten unmittelbar und unter Verwendung von Kern- oder Schlüsselementen aufeinander Bezug nehmen. Dabei handelt es sich im übrigen um eine allgemeine, nicht auf JA-ABER beschränkte Erscheinung. Das soll noch mit einem Beispiel aus einem anderen Zusammenhang belegt werden. Es geht um einen Einwurf eines zweiten Sprechers in die laufende Äußerung eines ersten Sprechers:

324 AN: aber der kleine handwerker is da * von der mentalität eben

325 AN: noch anders und glaubt alles selber machen zu müssen-

326 DE: **künstler**↓

¹² Wenn man dann, wissend dass der Kampf um Wörter zum politischen Kerngeschäft gehört, den Wortverteilungen weiter nachgeht, stellt man fest, dass *mannschaft* als Synonym von *team*, von ES, und nur von ES, im Gespräch 5mal verwendet wird. Aber solche Erweiterungen der lexikalischen Analyse sind ein späterer Schritt.

326 AN: ja **künstler**↓ genau↓ * aber * wir haben * trotz dieser

327 AN: →wir müssen die **künstler** sein die denen das

328 AN: vermitteln↓ also wir sind genauso gut **künstler**↓←

329 DE: ja

330 AN: wir haben **künstler** zu sein↓

331 DE: LACHT ja ja aber das=s ne

Das Wort *künstler* kommt im gesamten Gespräch sonst nicht vor. Aus der Betrachtung dieses Falles geht erneut hervor, dass die lokale Verdichtung mit den Strukturbedingungen und Verfahren der engen wechselseitigen Bezugnahme im Zusammenhang mit Aushandlungsprozessen zu tun hat. Das haben beide Kontexte gemeinsam.

5. Fazit

Die vorliegenden Ausführungen sind ein Plädoyer für die Weiterentwicklung von Recherche-Möglichkeiten für Interaktionskorpora. Dazu wurde ein experimentelles System auf der Grundlage von COSMAS II herangezogen, das in den 1990er Jahren im IDS entwickelt wurde. Die Ausführungen hier beziehen sich auf die in diesem System angelegten Möglichkeiten. Diese können im Prinzip natürlich auch in anderen Umgebungen funktionieren.

Das Zielobjekt ist ein komplexes Muster, JA-ABER, für das die Kombination von Zustimmung / Berücksichtigung / Einräumung und Widerspruch / oppositive Relevanzsetzung charakteristisch ist. Dieses Muster wurde gewählt, weil es einige Schwierigkeiten bietet, die für die Recherche von Gesprächseigenschaften typisch sind. Dazu gehören: eine große Variabilität der Ausdrucksweisen; interaktive Bezüge wie Sprecherwechsel, Folgen aufeinander reagierender Aktivitäten und Sequenzmuster; und schließlich relativ weit gespannte und in ihrer Ausdehnung nicht vorab begrenzbare Zusammenhänge. Als besonderes Problem erweist sich z. B. die Auffindung der Einräumungsteile und des strukturellen Zusammenhanges von JA und ABER.

Der Entwicklungsweg führte ausgehend von den Möglichkeiten der einfachen Suche von einzelnen Wortformen, die Erprobung von unterschiedlich großen Wortabständen zwischen den Zielausdrücken, die Zusammenstellung einer größeren Anzahl von Ausdrücken für eine multiple, gestaffelte Suche, über die Analyse von Formulierungsmustern, die für die Recherche geeignet erscheinen bis hin zu relativen Worthäufigkeiten in lokalen Kontexten.

Die angemessene Erfassung dieses Zusammenhanges von JA und ABER in komplexen Äußerungen erfordert eine Ausweitung der ursprünglich implementierten Rechercheverfahren. Das gilt für den Umgang mit Formulierungsmustern, die einerseits feste Strukturen, aber auch Variabilität aufweisen. Hierbei ist im Zweifelsfall eine größere Anzahl von Merkmalen abzu prüfen, was eine gestaffelte Suche erfordert. Und schließlich erfordert auch die Erfassung der relativen Häufigkeit von im Gesamtgespräch seltenen, nicht

vorab identifizierten Wörtern in begrenzten Kontexten eine Weiterentwicklung der ursprünglichen Möglichkeiten der COSMAS-II-Anwendung für Gesprächskorpora.

Der Versuch, die Möglichkeiten vorhandener und unter den gegebenen Voraussetzungen prinzipiell möglich erscheinender Recherche-Verfahren hat für sich genommen schon einen nicht zu unterschätzenden heuristischen Wert. Der von den Systemmöglichkeiten erzwungene sehr formale, viele interpretative Aspekte ausblendende und manchmal im Verhältnis zur gewohnten interpretativen Analysepraxis tangentielle oder auch „schräg“ wirkende Blick auf das Gesprächsmaterial schärft die Wahrnehmung der Eigenschaften des Materials in erstaunlicher Weise und stimuliert die Hypothesenbildung über strukturelle Eigenschaften des Zielobjekts.

Das Ziel der Entwicklung ist die Arbeit mit großen Korpora. Die damit verbundenen Vorteile werden teilweise auch schon bei der Arbeit mit einem noch begrenzten Auswahlkorpus deutlich. Es ergeben sich nicht nur große Fallmengen für die gesuchten Phänomene, sondern auch Einsichten in Häufigkeitsverteilungen. Die Arbeit mit Häufigkeitsverteilungen gehört nicht zur üblichen gesprächsanalytischen Praxis. Die gesprächsanalytische Orientierung auf die qualitative Untersuchung der den jeweiligen Fall konstituierenden Verfahren und Eigenschaften hat ein großes Erkenntnispotenzial, das auf keinen Fall zugunsten einer auf große Mengen von Material gestützten quantitativen Untersuchung aufgegeben werden soll. Aber die Arbeit mit größeren Korpora macht deutlich erkennbar, dass es zumindest heuristisch sehr fruchtbar ist, beide methodischen Perspektiven zu verbinden.¹³

Die Erweiterung der Musteranalyse um Häufigkeitsanalysen wird möglicherweise für die Entwicklung der qualitativen Analyse ein Schlüssel für weitere Entwicklungsschübe sein. Qualitative Analysen stützen sich in sehr vielen Fällen auf implizite Annahmen über Normalität und Auffälligkeit. Und diese Annahmen enthalten auch quantitative Größen. So z. B. wenn bestimmte Vorkommnisse am „Normalfall“ gemessen werden oder an dem, was „in der Regel“ passiert. Die Beurteilung von Auffälligkeit stützt sich einerseits auf die Behandlung des entsprechenden Aspekts in der Interaktion (z. B. wenn etwas als „erklärungsbedürftig“ markiert wird); aber spätestens die Analyse des Ausbleibens von Folgen (d. h. z. B. die Nichtbehandlung von möglicher Auffälligkeit) erfordert eine Bewertungsfolie („sonst“, „in der Regel“; oder auch nur „was auch hätte sein können“). Oft stützt sich die Analyse auf Vergleichsstellen im Kontext, die so interpretiert werden können, dass die fraglichen Normalformerwartungen an anderen Stellen beachtet werden bzw. aktivitätsleitend sind. Aber auch dabei liegen wiederum Annahmen zugrunde, nach denen diese Stellen signifikant sind. Für die Rekonstruktion des

¹³ Zur Debatte und die Verbindung von qualitativen und quantitativen Verfahren, auf die auch schon die Conversation Analysis reagiert hat (vgl. Schegloff 1993) und zu den Entwicklungsperspektiven vgl. u. a. Kallmeyer (2005).

interpretativen Charakters von Interaktion können Häufigkeitsuntersuchungen wichtige Anregungen geben. Solche Überlegungen haben am Beginn der Konversationsanalyse ja durchaus eine Rolle gespielt. Die Conversation Analysis war methodologisch gegen die übliche quantitative Sozialforschung gerichtet, und Harvey Sacks hat sich oft über die Sinnlosigkeit von Häufigkeitsberechnungen mit Stellen hinter dem Komma geäußert. Aber die relativ grobe Wahrnehmung von Verteilungen hat eine Rolle gespielt, z. B. wenn bei der Analyse der Sprecherwechselmechanismen festgestellt wird, dass etwas „in the vast majority of cases“ der Fall ist.¹⁴

Literatur

- Bodmer, Franck/Fach, Marcus L./Schmidt, Rudolf/Schütte, Wilfried (2002): Von der Tonbandaufnahme zur integrierten Text-Ton-Datenbank. Instrumente für die Arbeit mit Gesprächskorpora. In: Pusch/Raible (Hg.), S. 209–243.
- Cook, Guy (1995): Theoretical issues: transcribing the untranscribable. In: Leech/Myers/Thomas (eds.), S. 35–53.
- Fiehler, Reinhard/Barden, Birgit/Elstermann, Mechthild/Kraft, Barbara (2004): Eigenschaften gesprochener Sprache. (Studien zur Deutschen Sprache. Forschungen des Instituts für Deutsche Sprache Bd. 30). Tübingen.
- Kallmeyer, Werner (1997): Vom Nutzen des technologischen Wandels in der Sprachwissenschaft: Gesprächsanalyse und automatische Sprachverarbeitung. In: Zeitschrift für Literaturwissenschaft und Linguistik 107, S. 124–149.
- Kallmeyer, Werner (2000): Beraten und Betreuen. Zur gesprächsanalytischen Untersuchung von helfenden Interaktionen. In: Zeitschrift für Qualitative Bildungs-, Beratungs- und Sozialforschung 2, S. 227–252.
- Kallmeyer, Werner (2005): Qualitative Methoden. In: Ammon, Ulrich/Dittmar, Norbert/Mattheier, Klaus J./Trudgill, Peter (Hg.), Sociolinguistics/Soziolinguistik. An international Handbook of the Science of Language and Society/Ein internationales Handbuch zur Wissenschaft von Sprache und Gesellschaft. Bd. 1, Artikel 100. 2., überarbeitete Auflage. Berlin/New York. S. 978–992.
- Kallmeyer, Werner/Schmitt, Reinhold (1991a): Zur Markierung von oppositiven Reaktionen in komplexen Äußerungen. Mannheim, SFB 245 „Sprechen und Situation“.
- Kallmeyer, Werner/Schmitt, Reinhold (1991b): Zur Formulierungsdynamik von Äußerungen mit Fokusopposition. Mannheim, SFB 245 „Sprechen und Situation“.
- Kallmeyer, Werner/Schütte, Wilfried (2005): Der Umgang mit Gesprächskorpora am IDS. Die Recherche in der COSMAS-II-Gesprächsdatenbank. In: Schwitalla, Johannes/Wegstein, Werner (Hg.), Korpuslinguistik deutsch: synchron – diachron – kontrastiv. S. 71–80.
- Leech, Geoffrey N./Myers, Greg/Thomas, Jenny (eds.) (1995): Spoken English on computer: transcription, mark-up, and application. New York.
- Mair, Christian (2002): Tracking Ongoing Grammatical Change and Recent Diversification in Present-day Standard English. The Complementary Role of Small and Large Corpora. In: Pusch/Raible (Hg.), S. 9–26.
- Pasch, Renate/Brauß, Ursula/Breindl, Eva u. a. (2003): Handbuch der deutschen Konnektoren. Linguistische Grundlagen der Beschreibung und syntaktische Merkmale der

¹⁴ Vgl. Sacks/Schegloff/Jefferson (1974) und Schegloff (1968).

- deutschen Satzverknüpfen (Konjunktionen, Satzadverbien und Partikeln). (Schriften des Instituts für Deutsche Sprache, 9). Berlin/New York.
- Pusch, Claus D./Raible, Wolfgang (Hg.) (2002): Romanistische Korpuslinguistik: Korpora und gesprochene Sprache. Romance Corpus Linguistics: Corpora and Spoken Language. Tübingen (ScriptOralia 126).
- Sacks, Harvey/Schegloff, Emanuel/Jefferson, Gail (1974): A simplest systematics for the organization of turn taking for conversation. In: *Language* 50, S. 696–735.
- Schegloff, Emanuel (1993): Reflexions on Quantification in the Study of Conversation. In: *Research on Language and social Interaction*, 26, S. 99–128.
- Schmitz, Ulrich (2000): Neue Medien als Arbeitsinstrument der Linguistik. In: Kallmeyer, Werner (Hg.), *Sprache und neue Medien* (Institut für deutsche Sprache, Jahrbuch 1999). Berlin/New York. S. 253–274.
- Zifonun, Gisela/Hoffmann, Ludger/Strecker, Bruno et al. (1997): *Grammatik der deutschen Sprache*. (Schriften des Instituts für deutsche Sprache 7.17.3). Berlin/New York.

Anhang

A Korpusaufstellung

Die Aufstellung enthält die Wortmengen, die Vorkommen von ABER-Konnektoren, den prozentualen Anteil der ABER-Konnektoren zusammen an der Wortmenge und zu Vergleichszwecken Angaben zur Verwendung von NEIN in Voraussetzungskorrekturen und Zurückweisungen.

Telefonische Studentenberatung Dissertationskosten

| | |
|--------|-------|
| Wörter | 1428 |
| Aber | 3 |
| Nur | 0 |
| Bloß | 0 |
| Obwohl | 0 |
| Summe | 3 |
| % | 0,210 |
| Nein | 1 |

Medizinische Beratung Mongolismus

| | |
|--------|-------|
| Wörter | 2798 |
| Aber | 12 |
| Nur | 0 |
| Bloß | 0 |
| Obwohl | 0 |
| Summe | 12 |
| % | 0,429 |
| Nein | 0 |

Genetische Beratung

| | |
|--------|-------|
| Wörter | 780 |
| Aber | 6 |
| Nur | 0 |
| Bloß | 0 |
| Obwohl | 0 |
| Summe | 6 |
| % | 0,769 |
| Nein | 0 |

Schulpsychologische Beratung Erziehung

| | |
|--------|--------|
| Wörter | 12.976 |
| Aber | 57 |
| Nur | 2 |
| Bloß | 0 |
| Obwohl | 0 |
| Summe | 59 |
| % | 0,455 |
| Nein | 3 |

Studienberatung Theaterwissenschaftler

| | |
|--------|-------|
| Wörter | 995 |
| Aber | 7 |
| Nur | 0 |
| Bloß | 0 |
| Obwohl | 0 |
| Summe | 7 |
| % | 0,704 |
| Nein | 1 |

Studienberatung Luxemburgerin

| | |
|--------|-------|
| Wörter | 5128 |
| Aber | 50 |
| Nur | 3 |
| Bloß | 2 |
| Obwohl | 2 |
| Summe | 57 |
| % | 1,112 |
| Nein | 4 |

Studienberatung Student aus Mainz

| | |
|--------|-----|
| Wörter | 480 |
| Aber | 0 |
| Nur | 0 |
| Bloß | 0 |
| Obwohl | 0 |
| % | 0 |
| Nein | 1 |

Psychologische Beratung im Fernsehen Lämmle

| | |
|--------|-------|
| Wörter | 2763 |
| Aber | 19 |
| Nur | 1 |
| Bloß | 0 |
| Obwohl | 0 |
| Summer | 20 |
| % | 0,848 |
| Nein | 0 |

Nichtsesshaftenberatung Wiedereingliederung

| | |
|--------|-------|
| Wörter | 2357 |
| Aber | 14 |
| Nur | 3 |
| Bloß | 0 |
| Obwohl | 0 |
| Summe | 17 |
| % | 0,506 |
| Nein | 3 |

Finanzberatung Bausparen

| | |
|--------|-------|
| Wörter | 3612 |
| Aber | 24 |
| Nur | 5 |
| Bloß | 0 |
| Obwohl | 0 |
| Summe | 29 |
| % | 0,803 |
| Nein | 1 |

Referatbesprechung Professor und Student Sklavenreferat

| | |
|--------|-------|
| Wörter | 6580 |
| Aber | 40 |
| Nur | 2 |
| Bloß | 1 |
| Obwohl | 1 |
| Summe | 44 |
| % | 0,669 |
| Nein | 3 |

Prüfungsgespräch in der Universität

| | |
|--------|-------|
| Wörter | 6607 |
| Aber | 33 |
| Nur | 2 |
| Bloß | 0 |
| Obwohl | 0 |
| Summe | 35 |
| % | 0,530 |
| Nein | 1 |

Arbeitsbesprechung Cervantes

| | |
|--------|-------|
| Wörter | 8450 |
| Aber | 38 |
| Nur | 0 |
| Bloß | 0 |
| Obwohl | 0 |
| Summe | 38 |
| % | 0,450 |
| Nein | 1 |

Talkshow im Fernsehen Berlin 19 Aktionskünstler

| | |
|--------|-------|
| Wörter | 2771 |
| Aber | 12 |
| Nur | 0 |
| Bloß | 0 |
| Obwohl | 0 |
| Summe | 12 |
| % | 0,433 |
| Nein | 10 |

Talkshow im Fernsehen Feldbusch-Schwarzer

| | |
|--------|--------|
| Wörter | 17.667 |
| Aber | 129 |
| Nur | 15 |
| Bloß | 0 |
| Obwohl | 0 |
| Summe | 144 |
| % | 0,815 |
| Nein | 65 |

Politische Talkshow im Fernsehen Christiansen-Stoiber

| | |
|--------|--------|
| Wörter | 12.237 |
| Aber | 76 |
| Nur | 5 |
| Bloß | 0 |
| Obwohl | 0 |
| Summe | 81 |
| % | 0,662 |
| Nein | 27 |

B Suchoperatoren der COSMAS-Gesprächsrecherche

In der COSMAS-Gesprächsrecherche stehen folgende Operatoren zur Verfügung (vgl. auch Bodme Fach/Schmidt/Schütte 2002 und Kallmeyer/Schütte 2005):

- Der Wort-Operator *WORT(x)* sucht nach einem oder mehreren alternativen Suchbegriffen, die mit dem Platzhalteroperator * (Asterisk) versehen sein können. So findet *WORT(indirekt*)* Vorkommnisse von *indirekt*, *indirekte*, *indirekter*, *indirektes*, etc. Falls die Großschreibung in den Transkripten verwendet wird, kann durch entsprechende Optionen gesteuert werden, ob die Groß-/Kleinschreibung berücksichtigt werden soll. Durch die Trennung von Text und Annotationen in COSMAS II ist dieser Operator im Stande, Wörter zurückzuliefern, welche durch die Diskursgliederung oder Annotationen fragmentiert sind. Die Suchanfrage *WORT(indirekt)* findet demnach auch Textstellen wie *indi|rekt* mit einer Überlappungsgrenze, *indi*rekt* mit Mikropause oder *indirekt↓* mit Intonationsmarkierung.
- Operatoren für die Suche nach Wörtern mit prosodischen Merkmalen: So ist *BETONUNG* ein Suchoperator, der Wörter findet, bei welchen auffallende Betonungen markiert wurden, z. B. *i*ndirekt* mit auffallend betonter Silbe *in*. Die Operatoren *DEHNUNG* und *DEHNUNG-LANG* finden Wörter mit einer auffallenden Dehnung. *INTONATION(x)* sucht nach Wörtern mit aufsteigender, absteigender oder schwebender Intonation; z. B. liefert *INTONATION(fallend)* Textstellen wie *indirekt↓*.

- Mit dem Operator *GLEICH*(*X*, *Y*) lassen sich Wortformen und prosodische Merkmale kombinieren, z. B. das Wort *nicht* sowohl mit steigender Intonation als auch mit einer besonderen Dehnung. Als Treffer liefert das Verfahren ein Textstück wie *willst du nicht↑*, übergeht jedoch Fälle wie *glaubst du nicht↑*.
- Operatoren für Pausen verschiedener Länge: *PAUSE* für Pausen beliebiger Dauer, *PAUSE-KURZ* für Mikropausen (in unseren Diskurstranskripten ca. $\frac{1}{2}$ Sek.) und *PAUSE-LANG* für Pausen längerer Dauer ab 1 Sek. Diese drei Pausen-Operatoren lassen sich grundsätzlich je nach Art und Genauigkeit der in den Transkripten vorkommenden Pausenkodierungen beliebig erweitern.
- Operatoren für nicht-lexikalisierte Äußerungen: Nicht-lexikalisierte Äußerungen bilden in unseren Transkripten eine offene Menge von kurzen Beschreibungen in der Sprecherzeile, z. B. *und als ich sah dass er lacht LACHT KURZ sachte ich*. Diese Beschreibungen sind mit Großbuchstaben festgehalten, damit der Leser sie als Annotation erkennen kann. In COSMAS II werden sie in den Annotationsstrukturen gespeichert und über eine Reihe von speziellen Suchoperatoren recherchierbar gemacht, z. B. über den *LACHT*-Operator, der Beschreibungen wie *LACHT*, *LACHT KURZ*, *LACHANSATZ*, etc. wiederfindet. Da diese Beschreibungen nicht ausschließlich geschlossene Listen von Ausdrücken verwenden, sondern auch aus freiem Text bestehen können, haben wir einen Operator *VOCAL*(*x*) eingeführt, der mit freiem Text als Eingabe arbeitet, z. B. *VOCAL*(*LACHANSATZ*), *VOCAL*(*GRINST*), *VOCAL*(*SEUFZT*).
- Operatoren für die Diskursgliederung liefern Wörter zurück, die sich an markanten Stellen der Diskursgliederung wie Äußerungsgrenzen oder Simultanpassagen befinden. So liefern *ÄÜßER-BEG* und *ÄÜßER-END* jeweils das erste bzw. letzte Wort einer Äußerung zurück und *SIM-BEG* bzw. *SIM-END* jeweils das erste bzw. letzte Wort einer Simultanpassage.
- Operatoren für Sprecherdaten wie Altersangabe oder Geschlecht, die mit den anderen Operatoren kombinierbar sind. *IN-SP-GESCHL*(*X* *Y*) filtert die Suchanfrage *X* auf Sprecher mit Geschlecht *Y*; *IN-SP-ALTER*(*X* *Y*) filtert die Suchanfrage *X* auf Sprecher mit Alter *Y*. Bislang enthalten nur wenige Gesprächstranskripte diese Informationen, COSMAS II ist aber darauf vorbereitet, diese Daten bei Suchanfragen berücksichtigen zu können.
- Ein sprecherbezogener Wortabstandsoperator *ABSTAND*(*X* *nw* *Y*) gewährleistet, dass die gefundenen *X*- und *Y*-Textstellen vom selben Sprecher stammen und höchstens den angegebenen Abstand (in Worteinheiten) aufweisen. Dieser Operator kann *X*- und *Y*-Stellen miteinander kombinieren, zwischen denen Einschübe von anderen Sprechern vorkommen. So liefert *ABSTAND*(*WORT*(*nicht*) 3w *WORT*(*weil*)) *nicht-weil*-Wortpaare eines Sprechers zurück, die höchstens 3 Wörter auseinander liegen; *ABSTAND*(*WORT*(*nicht*) 0w *ÄÜßER-END*) liefert *nicht*-Stellen am Ende einer Äußerung zurück; und: *ABSTAND*(*WORT*(*nicht*) +1w *INTONATION*(*steigend*)) liefert *nicht*-Stellen zurück, die unmittelbar vor einem beliebigen Wort (+1w) mit steigender Intonation gesprochen wurden.
- Der Abfolgeoperator *ABSTAND*(*X* *ns* *Y*) betrachtet das Transkript als Abfolge von Zeitsegmenten. Eine Simultanpassage gilt in dieser Betrachtungsweise als ein Zeitsegment. Wörter außerhalb von Simultanpassagen bilden jedes für sich ein Segment. Im Gegensatz zum Wortabstand stellt dieser Abstandsoperator sowohl für Wörter eines einzelnen Sprechers als auch verschiedener Sprecher einen Bezug her. Er ist insbesondere als Instrument zur Untersuchung von Phänomenen in und um Simultanpassagen wichtig. Mit dem Abfolgeoperator können nicht-lexikalisierte Äußerungen berücksichtigt werden, die der Wortabstandsoperator außer Acht lässt. So sucht die komplexe Anfrage *ABSTAND*(*IN-SP-GESCHL*(*WORT*(*weil*) *fem*) +1s *IN-SP-GESCHL*(*WORT*(*nein*

nee nö) masc) nach Vorkommnissen von *weil* in Äußerungen von Sprecherinnen, denen unmittelbar (+ls) danach ein *nein* (oder *nee* oder *nö*) eines männlichen Teilnehmers folgt.

Transkriptionskonvention

| | | |
|--------------|---------------|--|
| ja aber | | simultane Äußerungen stehen übereinander; Anfang und |
| nein nie | mals | Ende werden auf den jeweiligen Textzeilen markiert |
| * | | kurze Pause (bis max. $\frac{1}{2}$ Sekunde) |
| ** | | etwas längere Pause (bis max. 1 Sekunde) |
| = | | Verschleifung (Elision) eines oder mehrerer Laute |
| | | <u>zwischen</u> Wörtern (z. B. <i>sa=mer</i> für <i>sagen wir</i>) |
| / | | Wortabbruch |
| ↑ | | steigende Intonation (z. B. <i>kommst du mit</i> ↑) |
| ↓ | | fallende Intonation (z. B. <i>jetzt stimmt es</i> ↓) |
| – | | schwebende Intonation (z. B. <i>ich sehe hier</i> –) |
| " | | auffällige Betonung (z. B. <i>aber ge"rn</i>) |
| : | | auffällige Dehnung (z. B. <i>ich war so: fertig</i>) |
| ←immer ich→ | | langsamer (relativ zum Kontext) |
| →immerhin← | | schneller (relativ zum Kontext) |
| >vielleicht< | | leiser (relativ zum Kontext) |
| <manchmal> | | lauter (relativ zum Kontext) |
| LACHT | | Wiedergabe nichtmorphemisierter Äußerung auf der Sprecherzeile in Großbuchstaben |
| K | IRONISCH | Kommentar zur Äußerung (auf der Kommentarzeile) |
| K& | QUIETSCHEN | nicht-kommunikatives (akustisches) Ereignis in der Gesprächssituation (auf der global. Kommentarzeile) |
| Ü | <i>Mutter</i> | Übersetzungszeile |